Enterotype detemine Notebook

Code

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Hide

```
library(ggplot2)
# library(export)
library(cluster)
library(clusterSim)
library(ade4)
library(pROC)
library(reshape2)
library(ggpubr)
library(coin)
library(gridExtra)
library(ape)
```

Hide

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()
```

```
[1] "C:/work/fmt_enterotype/a_microbiome/analysis"
```

Hide

```
source("pre_processing.R")
```

```
[1] 0.01
[1] 0.01
[1] 0.01
[1] 0.01
```

Hide

```
diff_entro_names <- function(t_pre_data_remove_c, pre_data.prj, pre_data.cluster){
    library(coin)
    quan<-seq(0.1, 0.9, 0.05)
    group <- pre_data.cluster
    prj <- pre_data.prj

    abundan_diff <- apply(t_pre_data_remove_c[,1:(ncol(t_pre_data_remove_c)-2)], 2, function(x)
{
        pt <-  as.data.frame(cbind(as.numeric(x), group, prj))
        colnames(pt)<-c('nx', 'group', 'prj')
        upt <- pt

        prj_list <- upt$prj
        list <- NULL
        for(i in unique(prj_list)){
            if (length(prj_list[prj_list %in% i]) > 2){
                list <- c(list, i)
            }
        }
        upt <- upt[upt$prj %in% list,]
        upt$nx <- as.numeric(as.character(upt$nx))
        tmp_test <- wilcox_test(nx ~ group | prj, upt)
        # tmp_test <- wilcox_test(nx ~ group , upt)
        pval_a <- NA
        pval_a <- pvalue(tmp_test)
        a_nx <- upt[upt$group %in% c('before1'), 'nx']
        p_nx <- upt[upt$group %in% c('before2'), 'nx']
        case <- mean(log(a_nx + 0.0001), trim = 0)
        # control <- quantile(log10(p_nx+ 0.0001), quan)
        control <- mean(log(p_nx+ 0.0001), trim = 0)
        gfc <- sum((case - control))/length(quan)
        return(c(pval_a, gfc))
    })
    abundan_diff_t <- data.frame(t(abundan_diff))
    colnames(abundan_diff_t) <- c('pval', 'gfc')
    abundan_diff_t$id <- rownames(abundan_diff_t)
    abundan_diff_t <- abundan_diff_t[!is.na(abundan_diff_t$pval),]
    abundan_diff_t$pval <- as.numeric(as.character(abundan_diff_t$pval))
    diff_qvalue <- p.adjust(abundan_diff_t$pval, method='fdr')
    abundan_diff_t <- cbind(abundan_diff_t, diff_qvalue)
    abundan_diff_t
}
```

Hide

```
## Enterotype in CDI
```

Hide

```
pre_data <- L6_rela_fil_sAg_others[,unique(c(meta_fil_config[meta_fil_config$Diease1 %in% c('CDI'), c('Previous_sra')]))]/1000
# head(colSums(pre_data))

pre_data_remove = noise.removal(pre_data, percent=0.01)
pre_data.dist=dist.JSD(pre_data_remove)

require(clusterSim)

pre_nclusters=NULL
for (k in 1:10) {
  if (k==1) {
    pre_nclusters[k]=NA
  } else {
    pre_data.cluster_temp=pam.clustering(pre_data.dist, k)
    pre_nclusters[k]=index.G1(t(pre_data_remove), pre_data.cluster_temp,  d = pre_data.dist,
                         centrotypes = "centroids")#medoids
  }
}
```

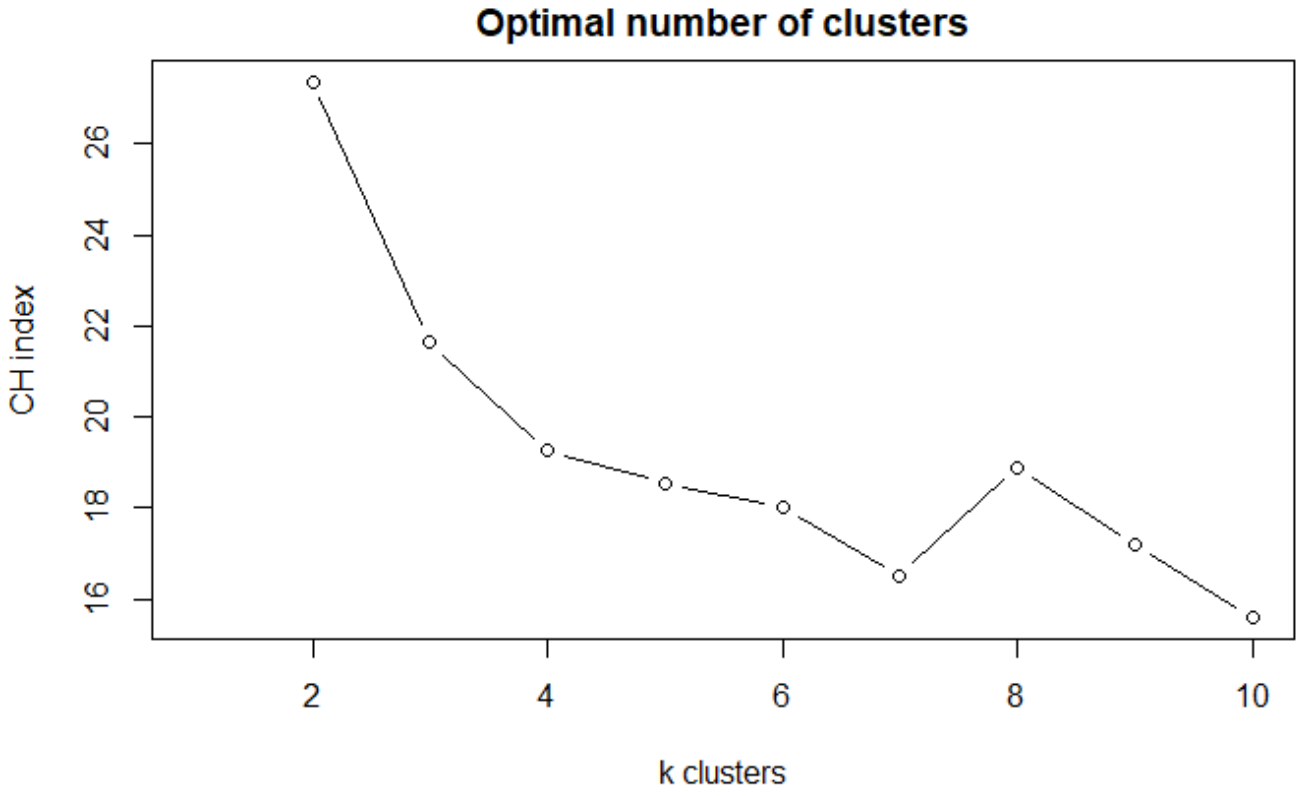
Hide

```
plot(pre_nclusters, type="b", xlab="k clusters", ylab="CH index",main="Optimal number of clusters")
pdf(file='./figure1/1main_CDI_class.pdf')
```

Hide

```
plot(pre_nclusters, type="b", xlab="Number of clusters", ylab="CH index")
dev.off()
```

```
png
  2
```



Hide

```
pre_data.cluster=pam.clustering(pre_data.dist, k=2)

pre_obs.silhouette=mean(silhouette(pre_data.cluster, pre_data.dist)[,3])
cat(pre_obs.silhouette) #0.1899451
```

```
0.1608902
```

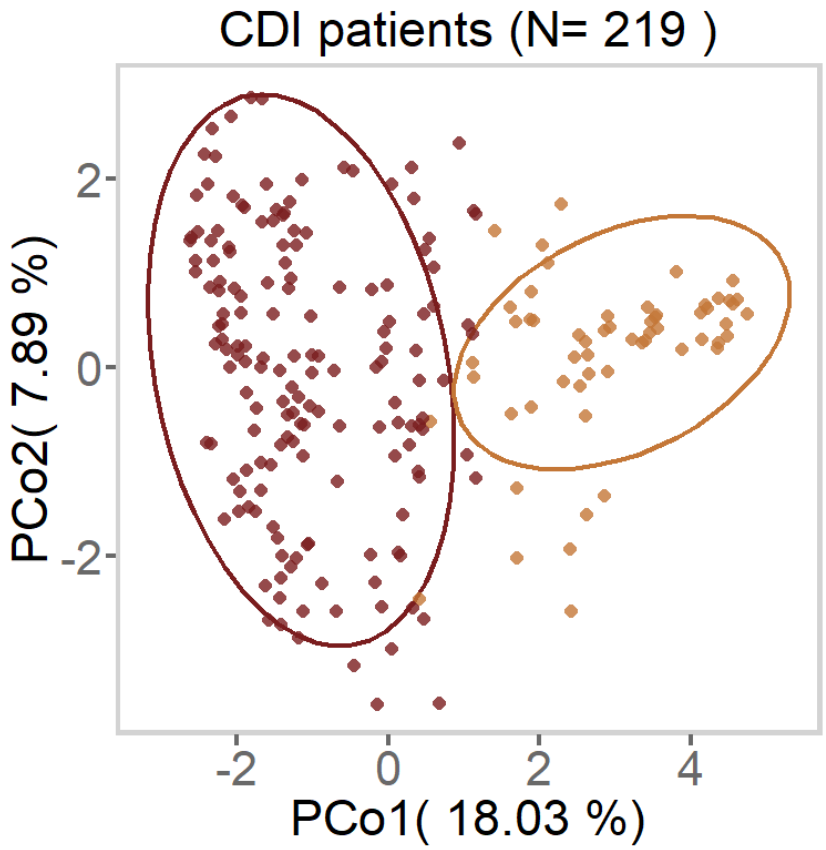
Hide

```
pre_obs.pcoa=dudi.pco(pre_data.dist, scannf=F, nf=2)


##pcoa

PCo1 <- pre_obs.pcoa$li[ ,1]
PCo2 = pre_obs.pcoa$li[ ,2]

library(ggplot2)
library(vegan)


Groupn<-'postfmt'
sample.groups <- 1


adonis(pre_data.dist ~ pre_data.cluster, permutations = 999)
```

```
Call:
adonis(formula = pre_data.dist ~ pre_data.cluster, permutations = 999)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                  Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
pre_data.cluster   1     693.2  693.16   34.28 0.13642  0.001 ***
Residuals        217    4387.8   20.22         0.86358
Total            218    5081.0                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
plotdata <- data.frame(rownames(pre_obs.pcoa$li),PCo1,PCo2, sample.groups, pre_data.cluster)
colnames(plotdata) <-c("sample","PCo1","PCo2","group", "data.cluster")
pc1 <-floor(pre_obs.pcoa$eig[1]*10000/sum(pre_obs.pcoa$eig))/100
pc2 <-floor(pre_obs.pcoa$eig[2]*10000/sum(pre_obs.pcoa$eig))/100


p<-ggplot(plotdata, alpha=I(0.8))+
  theme_classic()+
  stat_ellipse(aes(x=PCo1, y=PCo2, colour=as.factor(data.cluster),
                   group=as.factor(data.cluster)), level=0.9, size=1.5, show.legend = NA)+
  labs(title=paste("CDI patients (N=", length(PCo1),")"), x=paste("PCo1(",pc1,"%)"),y=paste("PC
o2(",pc2,"%)") , colour="Cluster")+
  geom_point(aes(x=PCo1,y=PCo2, colour=as.factor(data.cluster), shape=factor(sample.groups),
                 alpha = factor(sample.groups)),size=4)+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34
), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=0.95, legend.position = c(4, .65), legend.background=element_rect(fill = N
A), legend.text = element_text(size=18))+
  scale_colour_manual(values=c("#7A1D1E", "#C47737", "#E7A600"))+
  scale_alpha_manual(values = c(0.8))+
  theme(panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks = element_line(size=1.5, color
 ='dimgray'), axis.ticks.length = unit(7, "pt"));p
```



Hide

```
figli<-1
ggsave(paste("./figure1/1main_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1


beofre_entro_cdi <- cbind(rownames(pre_obs.pcoa$li), paste(rep('before', length(pre_data.cluste
r)), pre_data.cluster, sep = ''), meta_fil_config[match(rownames(pre_obs.pcoa$li), meta_fil_con
fig$Previous_sra), 'PRJ'])
```

Hide

```
library(coin)
pre_data.cluster <- beofre_entro_cdi[,2]
pre_data.prj <- beofre_entro_cdi[,3]

pre_simp_name <- sapply(as.character(rownames(pre_data_remove)), simp_names)
pre_data_remove_c <- rbind(pre_data_remove/100 + 1e-5, pre_data.cluster, pre_data.prj)
rownames(pre_data_remove_c) <- c(pre_simp_name, 'pre_data.cluster', 'pre_data.prj')
t_pre_data_remove_c<-t(pre_data_remove_c)
t_pre_data_remove_c <- data.frame(t_pre_data_remove_c, stringsAsFactors = F)
t_pre_data_remove_c$pre_data.cluster <- as.factor(as.character(t_pre_data_remove_c$pre_data.clu
ster))
t_pre_data_remove_c$pre_data.prj <- as.factor(t_pre_data_remove_c$pre_data.prj)
```

Hide

```
t_pre_data_remove_c$Enterobacteriaceae_ <- as.numeric(t_pre_data_remove_c$Enterobacteriaceae_)
wilcox_test(Enterobacteriaceae_ ~ pre_data.cluster| pre_data.prj, t_pre_data_remove_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Enterobacteriaceae_ by pre_data.cluster (before1, before2)
     stratified by pre_data.prj
Z = 9.2016, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

Hide

```
t_pre_data_remove_c$Bacteroides <- as.numeric(t_pre_data_remove_c$Bacteroides)
wilcox_test(Bacteroides ~ pre_data.cluster| pre_data.prj, t_pre_data_remove_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Bacteroides by pre_data.cluster (before1, before2)
     stratified by pre_data.prj
Z = -7.8029, p-value = 6.052e-15
alternative hypothesis: true mu is not equal to 0
```

Hide

```
t_pre_data_remove_c$Salmonella <- as.numeric(t_pre_data_remove_c$Salmonella)
wilcox_test(Salmonella ~ pre_data.cluster| pre_data.prj, t_pre_data_remove_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Salmonella by pre_data.cluster (before1, before2)
     stratified by pre_data.prj
Z = 5.4949, p-value = 3.909e-08
alternative hypothesis: true mu is not equal to 0
```

Hide

```
##sum genus to family for Enterobacteriaceae
keyword = 'f__Enterobacteriaceae'
genus_key <- sapply(as.character(grep(keyword, rownames(pre_data_remove), value = TRUE)), simp_
names)
t_pre_data_remove_c_Entero <- colSums(apply(t_pre_data_remove_c[,genus_key], 1, as.numeric))
names(t_pre_data_remove_c_Entero) == rownames(t_pre_data_remove_c)
```

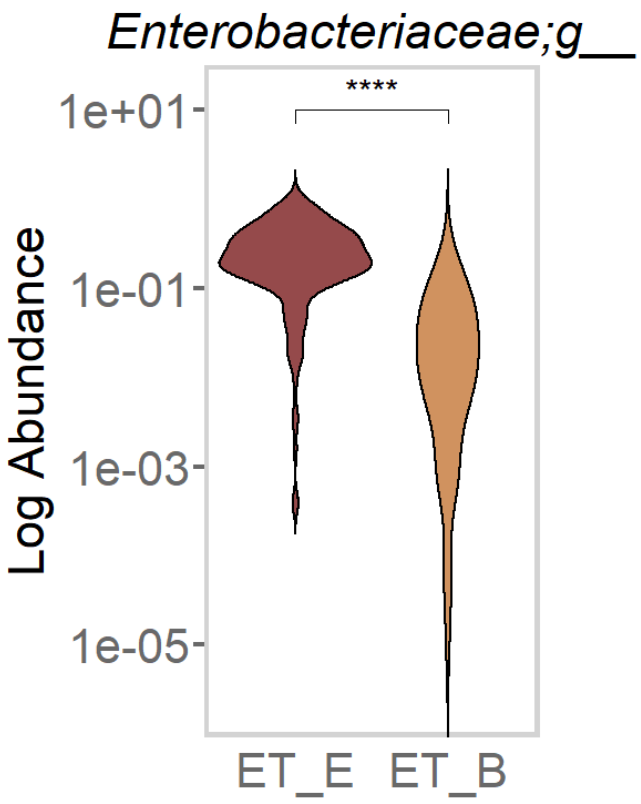
```
  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [26] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [51] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[101] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[126] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[151] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[176] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[201] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE
```

Hide

```
plot_pre_data <- cbind((t_pre_data_remove_c_Entero), t_pre_data_remove_c[,c('pre_data.cluster',
'Enterobacteriaceae_', 'Bacteroides')])
colnames(plot_pre_data) <- c('Enterobacteriaceae', 'pre_data.cluster', 'Enterobacteriaceae_',
'Bacteroides')
```

Hide

```
plot_pre_data1<- plot_pre_data
plot_pre_data1$pre_data.cluster <- ifelse(plot_pre_data1$pre_data.cluster == 'before1', 'ET_E',
'ET_B')
plot_pre_data1$pre_data.cluster <- factor(plot_pre_data1$pre_data.cluster , levels=c('ET_E', 'E
T_B'))

p1<-ggviolin(plot_pre_data1, x="pre_data.cluster", y="Enterobacteriaceae_", fill = "pre_data.cl
uster",#fill = "", Enterobacteriaceae_
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size
 = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.
signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab("Log Abundance")+labs
(title='Enterobacteriaceae;g__')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = elem
ent_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme( aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element
_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, colo
r ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p1
```



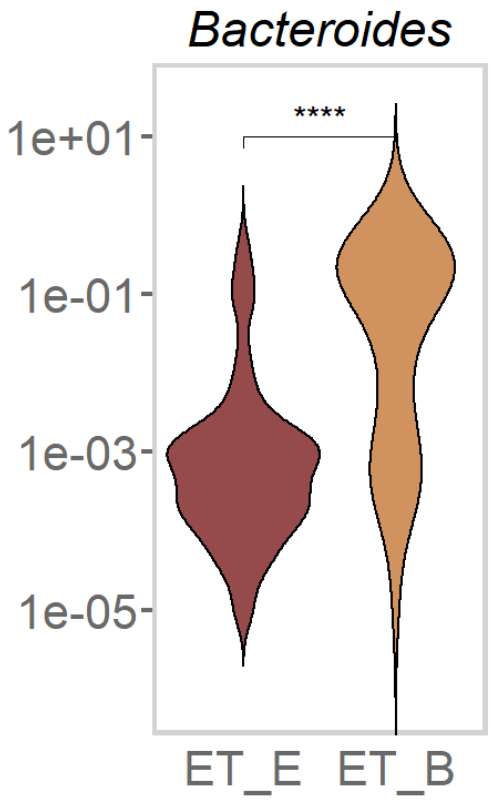
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p2<-ggviolin(plot_pre_data1, x="pre_data.cluster", y="Bacteroides", fill = "pre_data.cluster",#
fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size
 = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.
signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Bacte
roides')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = elem
ent_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element
_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, colo
r ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p2 #axis.title=element_text(size=21),
```



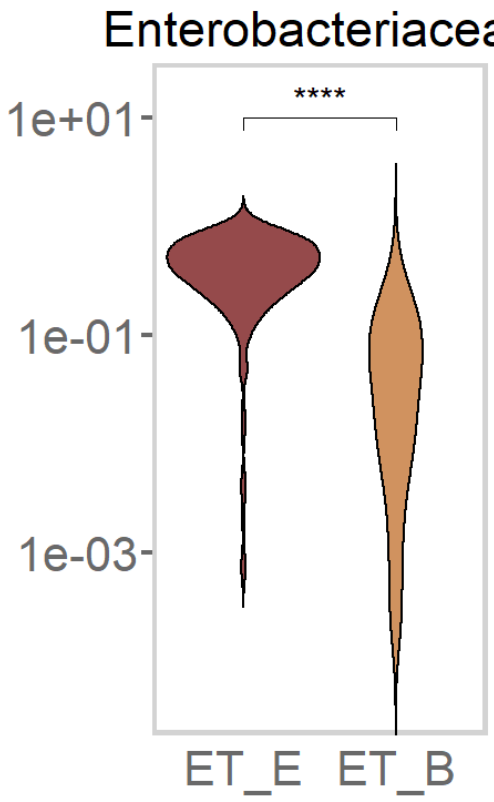
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p3<-ggviolin(plot_pre_data1, x="pre_data.cluster", y="Enterobacteriaceae", fill = "pre_data.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Enterobacteriaceae')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+#, face = 'italic'
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p3 #axis.title=element_text(size=21),
```



Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1
```

Hide

```
## Enterotype in IBD
```

Hide

```
pre_data1 <- L6_rela_fil_sAg_others[,unique(c(meta_fil_config[!meta_fil_config$Diease1 %in% c('CDI'), c('Previous_sra')]))]/1000
# head(colSums(pre_data1))

pre_data1_remove = noise.removal(pre_data1, percent=0.01)
pre_data1.dist=dist.JSD(pre_data1_remove)

require(clusterSim)

pre_nclusters=NULL
for (k in 1:10) {
  if (k==1) {
    pre_nclusters[k]=NA
  } else {
    pre_data1.cluster_temp=pam.clustering(pre_data1.dist, k)
    pre_nclusters[k]=index.G1(t(pre_data1_remove), pre_data1.cluster_temp,  d = pre_data1.dist,
                          centrotypes = "centroids")#medoids
  }
}
```

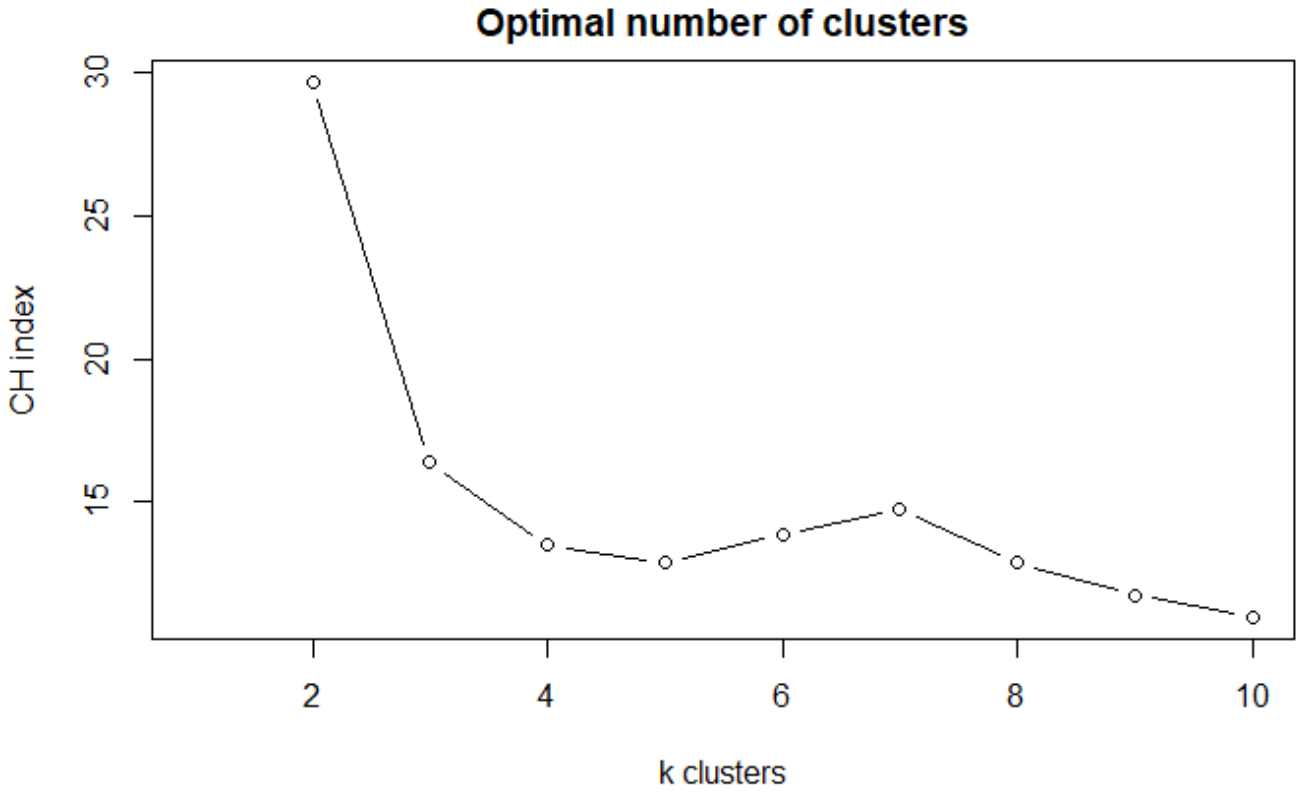
Hide

```
plot(pre_nclusters, type="b", xlab="k clusters", ylab="CH index",main="Optimal number of clusters")
pdf(file='./figure1/1main_IBD_class.pdf')
```

Hide

```
plot(pre_nclusters, type="b", xlab="Number of clusters", ylab="CH index")
dev.off()
```

```
png
  2
```



Hide

```
pre_data1.cluster=pam.clustering(pre_data1.dist, k=2)

pre_obs.silhouette=mean(silhouette(pre_data1.cluster, pre_data1.dist)[,3])
cat(pre_obs.silhouette) #0.1899451
```

```
0.1944136
```

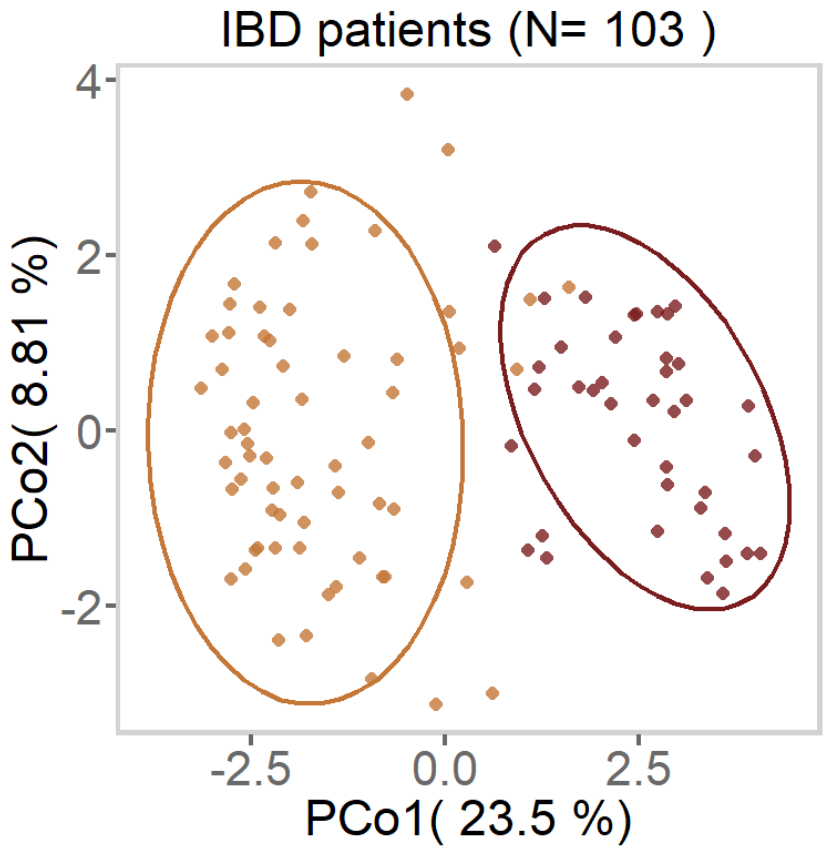
Hide

```
pre_obs.pcoa=dudi.pco(pre_data1.dist, scannf=F, nf=2)

##pcoa

PCo1 <- pre_obs.pcoa$li[ ,1]
PCo2 = pre_obs.pcoa$li[ ,2]

library(ggplot2)
library(vegan)

Groupn<-'postfmt'

sample.groups <- 1

pre_data1.cluster <- -1 * as.numeric(pre_data1.cluster) + 3

adonis(pre_data1.dist ~ pre_data1.cluster, permutations = 999)
```

```
Call:
adonis(formula = pre_data1.dist ~ pre_data1.cluster, permutations = 999)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                   Df SumsOfSqs MeanSqs F.Model     R2 Pr(>F)
pre_data1.cluster   1    420.06  420.06  23.185 0.1867  0.001 ***
Residuals         101   1829.88   18.12         0.8133
Total             102   2249.94                 1.0000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
plotdata <- data.frame(rownames(pre_obs.pcoa$li),PCo1,PCo2, sample.groups, pre_data1.cluster)
colnames(plotdata) <-c("sample","PCo1","PCo2","group", "data.cluster")
pc1 <-floor(pre_obs.pcoa$eig[1]*10000/sum(pre_obs.pcoa$eig))/100
pc2 <-floor(pre_obs.pcoa$eig[2]*10000/sum(pre_obs.pcoa$eig))/100


p<-ggplot(plotdata, alpha=I(0.8))+
  theme_classic()+
  stat_ellipse(aes(x=PCo1, y=PCo2, colour=as.factor(data.cluster),
                   group=as.factor(data.cluster)), level=0.9, size=1.5, show.legend = NA)+
  labs(title=paste("IBD patients (N=", length(PCo1),")"), x=paste("PCo1(",pc1,"%)"),y=paste("PC
o2(",pc2,"%)") , colour="Cluster")+
  geom_point(aes(x=PCo1,y=PCo2, colour=as.factor(data.cluster), shape=factor(sample.groups),
                 alpha = factor(sample.groups)),size=4)+
  # theme(plot.title = element_text(size=21, hjust = 0.5),
  #     # title=element_text(family ="sans", size=21),
  #       text=element_text(family ="sans", size=18),)+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34
), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=0.95, legend.position = c(4, .65), legend.background=element_rect(fill = N
A), legend.text = element_text(size=18))+
  scale_colour_manual(values=c("#7A1D1E", "#C47737", "#E7A600"))+
  scale_alpha_manual(values = c(0.8))+
  theme(panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks = element_line(size=1.5, color
 ='dimgray'), axis.ticks.length = unit(7, "pt"));p
```



Hide

```
# figli<-1
ggsave(paste("./figure1/1main_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

# grid.arrange(p1,p2,nrow=1)
beofre_entro_ibd <- cbind(rownames(pre_obs.pcoa$li), paste(rep('before', length(pre_data1.clust
er)), pre_data1.cluster, sep = ''), meta_fil_config[match(rownames(pre_obs.pcoa$li), meta_fil_c
onfig$Previous_sra), 'PRJ'])
```

Hide

```
library(coin)
pre_data1.cluster <- beofre_entro_ibd[,2]
pre_data1.prj <- beofre_entro_ibd[,3]

pre_simp_name <- sapply(as.character(rownames(pre_data1_remove)), simp_names)
pre_data1_remove_c <- rbind(pre_data1_remove/100 + 1e-5, pre_data1.cluster, pre_data1.prj)
rownames(pre_data1_remove_c) <- c(pre_simp_name, 'pre_data1.cluster', 'pre_data1.prj')
t_pre_data1_remove_c<-t(pre_data1_remove_c)
t_pre_data1_remove_c <- data.frame(t_pre_data1_remove_c, stringsAsFactors = F)
t_pre_data1_remove_c$pre_data1.cluster <- as.factor(as.character(t_pre_data1_remove_c$pre_data
1.cluster))
t_pre_data1_remove_c$pre_data1.prj <- as.factor(t_pre_data1_remove_c$pre_data1.prj)
```

Hide

```
t_pre_data1_remove_c$Enterobacteriaceae_ <- as.numeric(t_pre_data1_remove_c$Enterobacteriaceae
_)
wilcox_test(Enterobacteriaceae_ ~ pre_data1.cluster| pre_data1.prj, t_pre_data1_remove_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Enterobacteriaceae_ by
     pre_data1.cluster (before1, before2)
     stratified by pre_data1.prj
Z = 5.2115, p-value = 1.873e-07
alternative hypothesis: true mu is not equal to 0
```

Hide

```
t_pre_data1_remove_c$Bacteroides <- as.numeric(t_pre_data1_remove_c$Bacteroides)
wilcox_test(Bacteroides ~ pre_data1.cluster| pre_data1.prj, t_pre_data1_remove_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Bacteroides by
     pre_data1.cluster (before1, before2)
     stratified by pre_data1.prj
Z = -7.7851, p-value = 6.964e-15
alternative hypothesis: true mu is not equal to 0
```

Hide

```
##sum genus to family for Enterobacteriaceae
keyword = 'f__Enterobacteriaceae'
genus_key <- sapply(as.character(grep(keyword, rownames(pre_data1_remove), value = TRUE)), simp
_names)
t_pre_data1_remove_c_Entero <- colSums(apply(t_pre_data1_remove_c[,genus_key], 1, as.numeric))
names(t_pre_data1_remove_c_Entero) == rownames(t_pre_data1_remove_c)
```

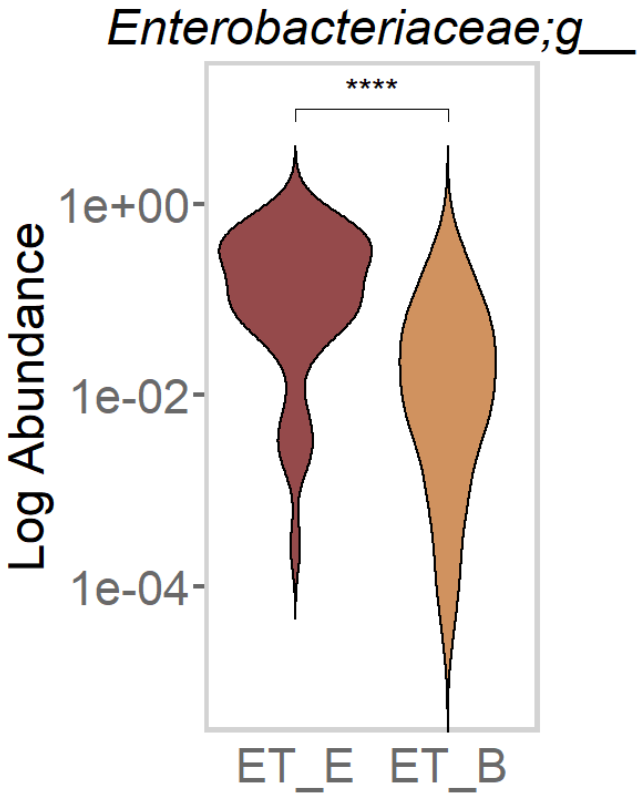
```
  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [26] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [51] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[101] TRUE TRUE TRUE
```

Hide

```
plot_pre_data1 <- cbind((t_pre_data1_remove_c_Entero), t_pre_data1_remove_c[,c('pre_data1.clust
er', 'Enterobacteriaceae_', 'Bacteroides')])
colnames(plot_pre_data1) <- c('Enterobacteriaceae', 'pre_data1.cluster', 'Enterobacteriaceae_',
'Bacteroides')
```

Hide

```
plot_pre_data11<- plot_pre_data1
plot_pre_data11$pre_data1.cluster <- ifelse(plot_pre_data11$pre_data1.cluster == 'before1', 'ET
_E', 'ET_B')
plot_pre_data11$pre_data1.cluster <- factor(plot_pre_data11$pre_data1.cluster , levels=c('ET_E'
, 'ET_B'))

p1<-ggviolin(plot_pre_data11, x="pre_data1.cluster", y="Enterobacteriaceae_", fill = "pre_data
1.cluster",#fill = "", Enterobacteriaceae_
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size
 = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.
signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab("Log Abundance")+labs
(title='Enterobacteriaceae;g__')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = elem
ent_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme( aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element
_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, colo
r ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p1
```



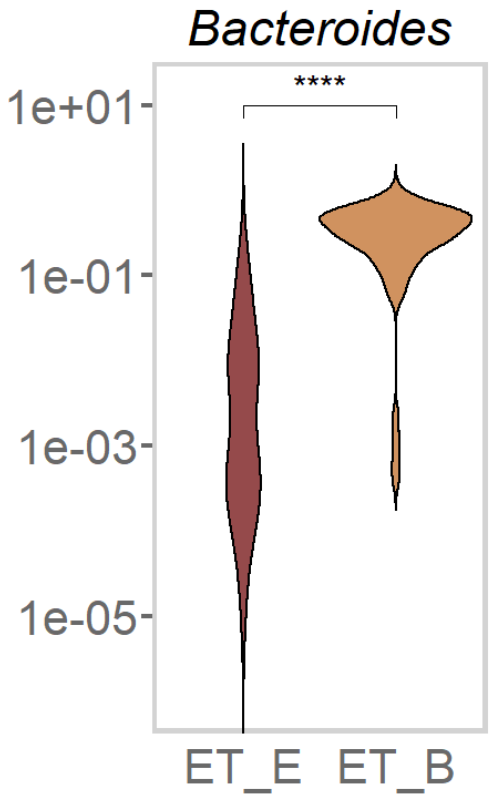
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p2<-ggviolin(plot_pre_data11, x="pre_data1.cluster", y="Bacteroides", fill = "pre_data1.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Bacteroides')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p2 #axis.title=element_text(size=21),
```



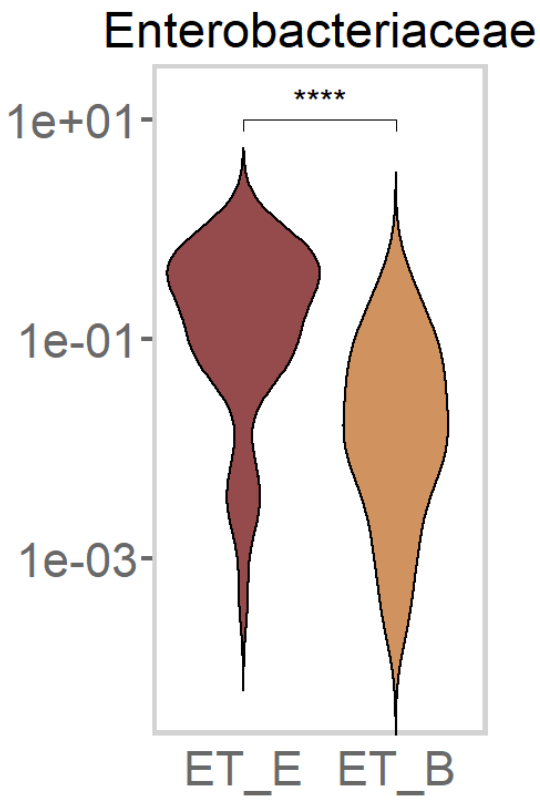
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p3<-ggviolin(plot_pre_data11, x="pre_data1.cluster", y="Enterobacteriaceae", fill = "pre_data1.
cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size
 = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.
signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Enter
obacteriaceae')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34
), axis.title.y = element_text(size=34), axis.ticks = element_blank())+#, face = 'italic'
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element
_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, colo
r ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p3 #axis.title=element_text(size=21),
```



Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1
```

Hide

```
## Enterotype in CDI & IBD
```

Hide

```
##before
pre_data0 <- L6_rela_fil_sAg_others[,unique(c(meta_fil_config$Previous_sra))]/1000

# head(colSums(pre_data0))

pre_data0_remove = noise.removal(pre_data0, percent=0.01)
pre_data0.dist=dist.JSD(pre_data0_remove)

require(clusterSim)

pre_nclusters=NULL
for (k in 1:10) {
  if (k==1) {
    pre_nclusters[k]=NA
  } else {
    pre_data0.cluster_temp=pam.clustering(pre_data0.dist, k)
    pre_nclusters[k]=index.G1(t(pre_data0_remove), pre_data0.cluster_temp,  d = pre_data0.dist,
                         centrotypes = "centroids")#medoids
  }
}
```

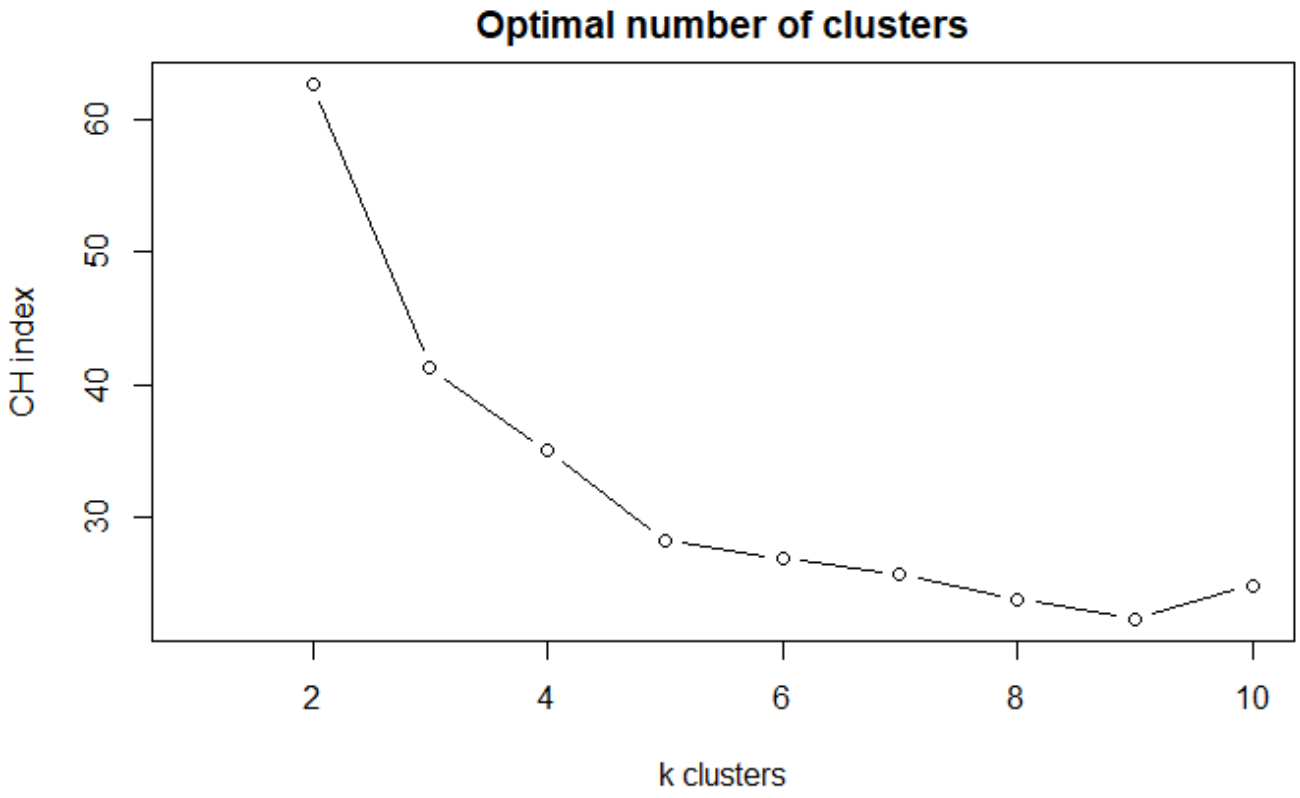
Hide

```
plot(pre_nclusters, type="b", xlab="k clusters", ylab="CH index",main="Optimal number of cluste
rs")
pdf(file='./figure1/1main_combine_class.pdf')
```

Hide

```
plot(pre_nclusters, type="b", xlab="Number of clusters", ylab="CH index")
dev.off()
```

```
png
  2
```



Hide

```
pre_data0.cluster=pam.clustering(pre_data0.dist, k=2)

pre_obs.silhouette=mean(silhouette(pre_data0.cluster, pre_data0.dist)[,3])
cat(pre_obs.silhouette) #0.1899451
```

```
0.1630546
```

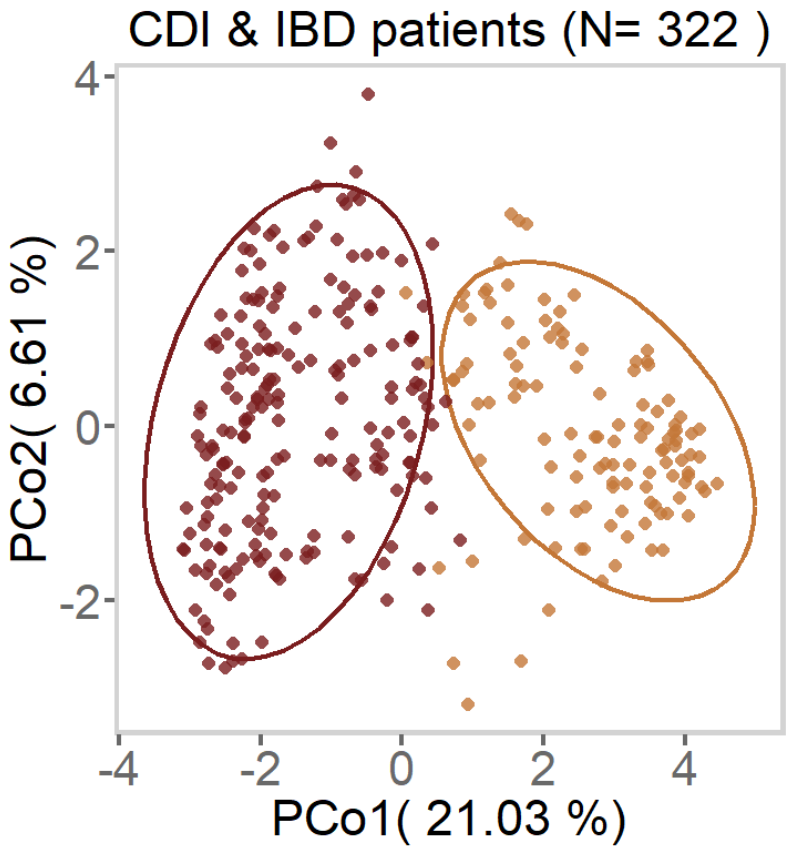
Hide

```
pre_obs.pcoa=dudi.pco(pre_data0.dist, scannf=F, nf=2)

##pcoa

PCo1 <- pre_obs.pcoa$li[ ,1]
PCo2 = pre_obs.pcoa$li[ ,2]

library(ggplot2)

Groupn<-'postfmt'

sample.groups <- 1

adonis(pre_data0.dist ~ pre_data0.cluster, permutations = 999)
```

```
Call:
adonis(formula = pre_data0.dist ~ pre_data0.cluster, permutations = 999)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

                  Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
pre_data0.cluster   1    1299.9  1299.9  64.354 0.16743  0.001 ***
Residuals         320    6464.0    20.2         0.83257
Total             321    7763.9                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
plotdata <- data.frame(rownames(pre_obs.pcoa$li),PCo1,PCo2, sample.groups, pre_data0.cluster)
colnames(plotdata) <-c("sample","PCo1","PCo2","group", "data.cluster")
pc1 <-floor(pre_obs.pcoa$eig[1]*10000/sum(pre_obs.pcoa$eig))/100
pc2 <-floor(pre_obs.pcoa$eig[2]*10000/sum(pre_obs.pcoa$eig))/100

p<-ggplot(plotdata, alpha=I(0.8))+
  theme_classic()+
  stat_ellipse(aes(x=PCo1, y=PCo2, colour=as.factor(data.cluster),
                   group=as.factor(data.cluster)), level=0.9, size=1.5, show.legend = NA)+
  labs(title=paste("CDI & IBD patients (N=", length(PCo1),")"), x=paste("PCo1(",pc1,"%)"),y=pas
te("PCo2(",pc2,"%)") , colour="Cluster")+
  geom_point(aes(x=PCo1,y=PCo2, colour=as.factor(data.cluster), shape=factor(sample.groups),
                 alpha = factor(sample.groups)),size=4)+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34
), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=1, legend.position = c(4, .65), legend.background=element_rect(fill = NA),
legend.text = element_text(size=18))+
  scale_colour_manual(values=c("#7A1D1E", "#C47737", "#E7A600"))+
  scale_alpha_manual(values = c(0.8))+
  theme(panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks = element_line(size=1.5, color
 ='dimgray'), axis.ticks.length = unit(7, "pt"));p
```



Hide

```
ggsave(paste("./figure1/1main_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1


beofre_entro <- cbind(rownames(pre_obs.pcoa$li), paste(rep('before', length(pre_data0.cluste
r)), pre_data0.cluster, sep = ''), meta_fil_config[match(rownames(pre_obs.pcoa$li), meta_fil_co
nfig$Previous_sra), 'PRJ'])
```

Hide

```
library(coin)
pre_data0.cluster <- beofre_entro[,2]
pre_data0.prj <- beofre_entro[,3]

# options(stringsAsFactors = FALSE)
pre_simp_name <- sapply(as.character(rownames(pre_data0_remove)), simp_names)
pre_data0_remove_dat <- data.frame(pre_data0_remove, stringsAsFactors = F)
pre_data0_remove_c <- as.data.frame(rbind(pre_data0_remove_dat/100 + 1e-5, pre_data0.cluster, p
re_data0.prj), stringsAsFactors =F)
rownames(pre_data0_remove_c) <- c(pre_simp_name, 'pre_data0.cluster', 'pre_data0.prj')
t_pre_data0_remove<-t(pre_data0_remove_c)
t_pre_data0_remove_c <- as.data.frame(t_pre_data0_remove, stringsAsFactors = F)

t_pre_data0_remove_c$pre_data0.cluster <- as.factor(as.character(t_pre_data0_remove_c$pre_data
0.cluster))
t_pre_data0_remove_c$pre_data0.prj <- as.factor(t_pre_data0_remove_c$pre_data0.prj)

t_pre_data0_remove_c$Enterobacteriaceae_ <- as.numeric(t_pre_data0_remove_c$Enterobacteriaceae
_)
wilcox_test(Enterobacteriaceae_ ~ pre_data0.cluster| pre_data0.prj, t_pre_data0_remove_c)
```

```
    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Enterobacteriaceae_ by
     pre_data0.cluster (before1, before2)
     stratified by pre_data0.prj
Z = 10.462, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

Hide

```
t_pre_data0_remove_c$Bacteroides <- as.numeric(t_pre_data0_remove_c$Bacteroides)
wilcox_test(Bacteroides ~ pre_data0.cluster| pre_data0.prj, t_pre_data0_remove_c)
```

```
    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Bacteroides by
     pre_data0.cluster (before1, before2)
     stratified by pre_data0.prj
Z = -10.999, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

Hide

```
##sum genus to family for Enterobacteriaceae
keyword = 'f__Enterobacteriaceae'
genus_key <- sapply(as.character(grep(keyword, rownames(pre_data0_remove), value = TRUE)), simp
_names)
t_pre_data0_remove_c_Entero <- colSums(apply(t_pre_data0_remove_c[,genus_key], 1, as.numeric))
names(t_pre_data0_remove_c_Entero) == rownames(t_pre_data0_remove_c)
```

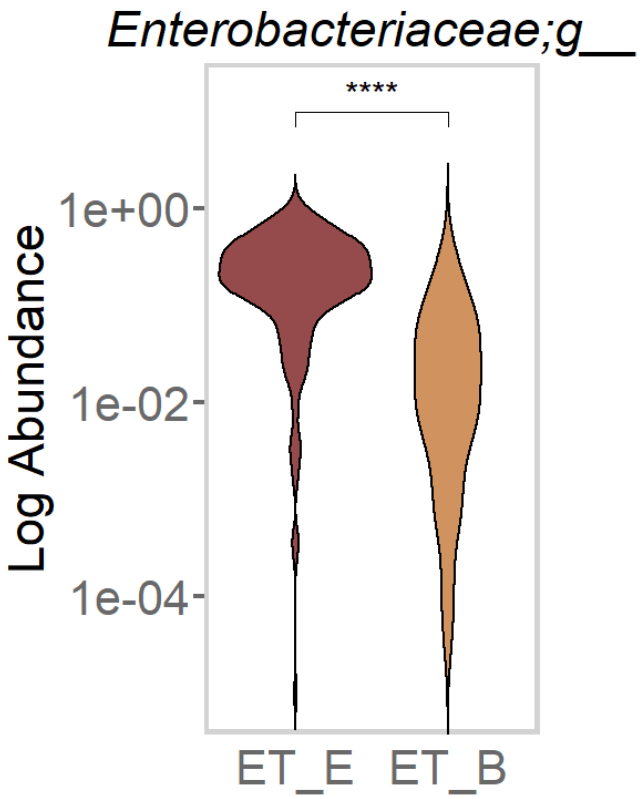
```
  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [26] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [51] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[101] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[126] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[151] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[176] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[201] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[226] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[251] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[276] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE
```

Hide

```
plot_pre_data0 <- cbind((t_pre_data0_remove_c_Entero), t_pre_data0_remove_c[,c('pre_data0.clust
er', 'Enterobacteriaceae_', 'Bacteroides')])
colnames(plot_pre_data0) <- c('Enterobacteriaceae', 'pre_data0.cluster', 'Enterobacteriaceae_',
'Bacteroides')
```

Hide

```
plot_pre_data01<- plot_pre_data0
plot_pre_data01$pre_data0.cluster <- ifelse(plot_pre_data01$pre_data0.cluster == 'before1', 'ET_E', 'ET_B')
plot_pre_data01$pre_data0.cluster <- factor(plot_pre_data01$pre_data0.cluster , levels=c('ET_E', 'ET_B'))

p1<-ggviolin(plot_pre_data01, x="pre_data0.cluster", y="Enterobacteriaceae_", fill = "pre_data0.cluster",#fill = "", Enterobacteriaceae_
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab("Log Abundance")+labs(title='Enterobacteriaceae;g__')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme( aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p1
```



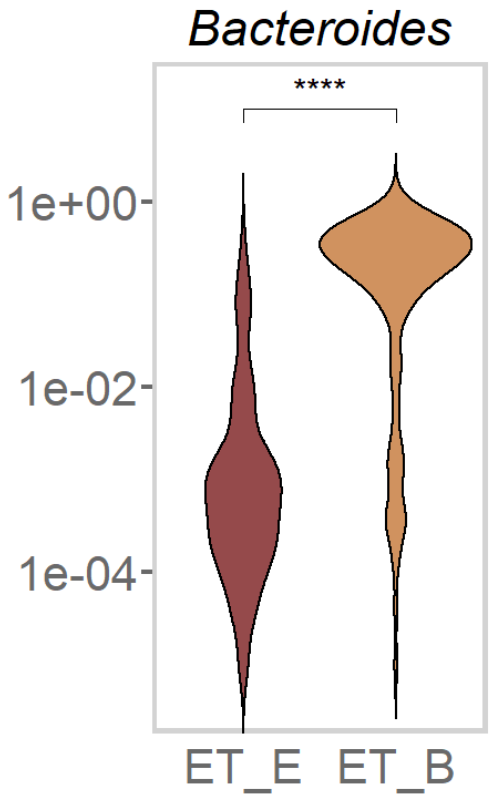
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p2<-ggviolin(plot_pre_data01, x="pre_data0.cluster", y="Bacteroides", fill = "pre_data0.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Bacteroides')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p2 #axis.title=element_text(size=21),
```



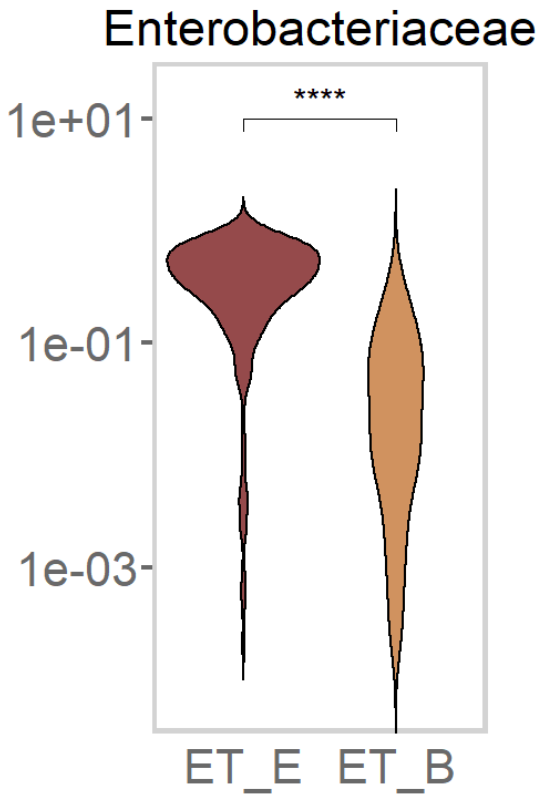
Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1

p3<-ggviolin(plot_pre_data01, x="pre_data0.cluster", y="Enterobacteriaceae", fill = "pre_data0.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Enterobacteriaceae')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+#, face = 'italic'
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p3 #axis.title=element_text(size=21),
```



Hide

```
ggsave(paste("./figure1/fig_s1_", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1
```

Hide

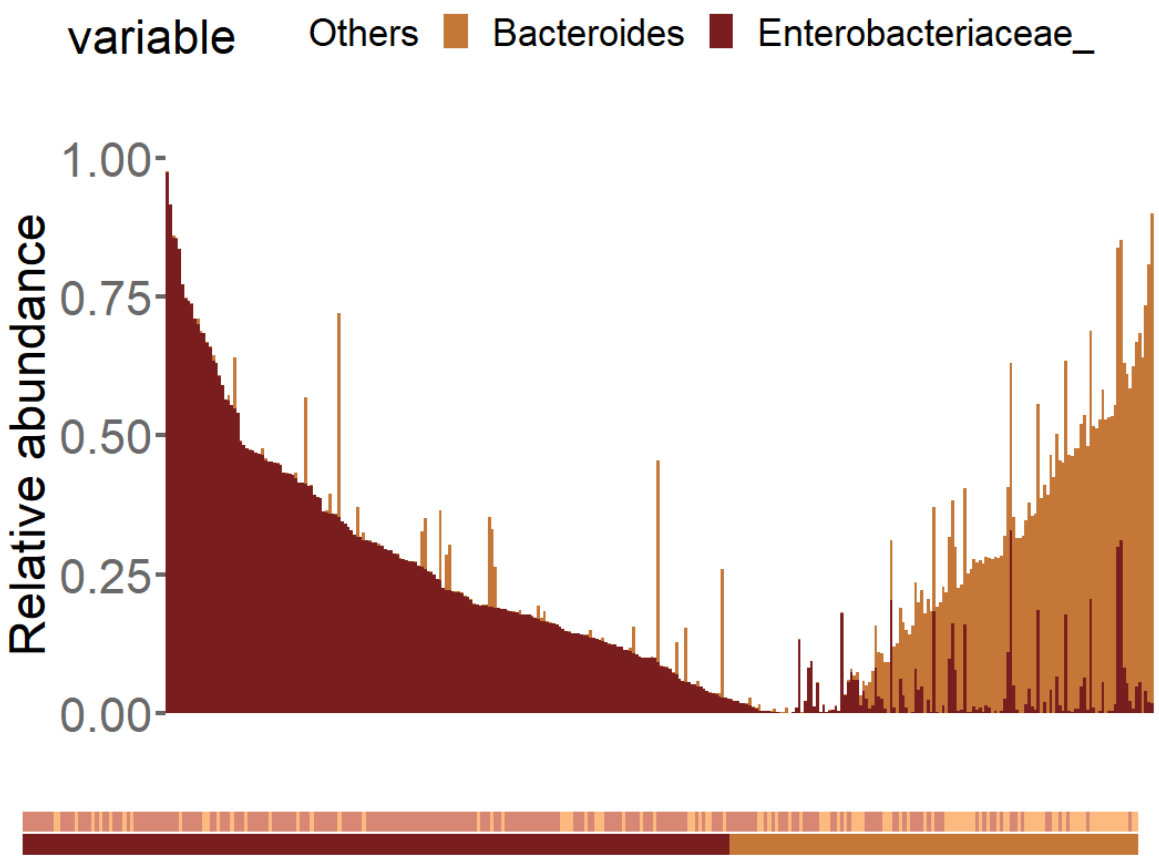
```
##dominat bar plot
plot_pre_data02 <- plot_pre_data0
plot_pre_data02$pre_data0.cluster <- ifelse(plot_pre_data02$pre_data0.cluster == 'before1', 'ET
_E', 'ET_B')
plot_pre_data02$pre_data0.cluster <- factor(plot_pre_data02$pre_data0.cluster , levels=c('ET_E'
, 'ET_B'))

plot_pre_data02$sample <- rownames(plot_pre_data02)
plot_pre_data02$Enterobacteriaceae_ <- as.numeric(plot_pre_data02$Enterobacteriaceae_)
plot_pre_data02$Bacteroides <- as.numeric(plot_pre_data02$Bacteroides)
plot_pre_data02$Others <- 1 - (plot_pre_data02$Enterobacteriaceae_ + plot_pre_data02$Bacteroide
s)

tmp = meta_fil_config1[,c('Previous_sra', 'Diease1')]
colnames(tmp) <- c('sample', 'Diease1')
tmp[tmp$Diease1 %in% c('CD', 'UC'), 'Diease1'] <- 'IBD'
plot_pre_data02 <- merge(plot_pre_data02, tmp, by='sample')


plot_pre_data02$pre_data0.cluster <- factor(plot_pre_data02$pre_data0.cluster, levels = c('ET_
E', 'ET_B'))
plot_pre_data02$Diease1 <- factor(plot_pre_data02$Diease1, levels = c('CDI', 'IBD'))

plot_pre_data02_eb <- plot_pre_data02[order((3*as.numeric(plot_pre_data02$pre_data0.cluster) +
ifelse(plot_pre_data02$pre_data0.cluster == 'ET_E', 1-as.numeric(plot_pre_data02$Enterobacteria
ceae_), as.numeric(plot_pre_data02$Bacteroides)))),] #as.numeric(plot_pre_data02$Diease1)


library(reshape2)
plot_pre_data02_eb_melt <- melt(plot_pre_data02_eb, measure.vars = c('Enterobacteriaceae_', 'Ba
cteroides', 'Others'))
plot_pre_data02_eb_melt$variable <- factor(plot_pre_data02_eb_melt$variable, levels = rev(c('En
terobacteriaceae_', 'Bacteroides', 'Others')))


plot_pre_data02_eb_melt$sample <- factor(plot_pre_data02_eb_melt$sample, levels = unique(plot_p
re_data02_eb_melt$sample))
```

Hide

```
p1<- ggplot(plot_pre_data02_eb_melt, aes(x=sample, y=value, fill=variable))+
  geom_bar(stat="identity", width = 1)+
  # geom_hline(yintercept = 0.05, size=1.5, linetype=2)+
  # geom_vline(xintercept = sum(t_pre_data0_remove_c_eb_melt$pre_data0.cluster == 'ET_E')/3, )+
  labs(x= c(''), y=c('Relative abundance'), title = c(''))+
    scale_fill_manual(values=c('white', "#C47737", "#7A1D1E", '#C6832A', '#962E2B',  '#4E86C6'
, '#E7A600', "#4D9127", "#90908D", "#962E2B", "#4E86C6", "#4D9127", "#90908D", 'lightgrey'))+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.text.x = element_blank(), axis.
title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks.y = element_l
ine(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"), axis.ticks=element_blank())
+
 theme(aspect.ratio = 0.618, legend.background=element_blank(), legend.position=c(1.75, 0.6)
       ,panel.background = element_rect(fill = NA, colour = "NA", size = 3)#lightgrey
       ,axis.line=element_line(colour="NA")
       ,legend.key = element_rect(fill = NA, color = NA), panel.border = element_blank(), pane
l.grid = element_blank())+
   guides(colour = guide_legend(override.aes = list(size=5)));
```

Hide

```
p2<-ggplot(plot_pre_data02_eb_melt)+
    geom_bar(mapping = aes(x = sample, y = 1, fill = Diease1),
       stat = "identity",
       width = 1)+
      theme_void()+
    scale_fill_manual(values=c("#D78774", "#FCBA7F"))+
      theme(panel.spacing.x = unit(1, "mm"), aspect.ratio = 0.02);

p3<-ggplot(plot_pre_data02_eb_melt)+
    geom_bar(mapping = aes(x = sample, y = 1, fill = pre_data0.cluster),
       stat = "identity",
       width = 1)+
      theme_void()+
    scale_fill_manual(values=c("#7A1D1E", "#C47737"))+
      theme(panel.spacing.x = unit(1, "mm"), aspect.ratio = 0.02)

ggpubr::ggarrange(p1, p2, p3, ncol = 1, nrow = 3, common.legend = TRUE, heights = c(1, 0.03, 0.
03))
```



Hide

```
ggsave("./figure1/1main_two_dominate_genus.pdf", device = 'pdf')
```

```
Saving 12.9 x 8 in image
```

Hide

```
##DMM
library(DirichletMultinomial)
full <- FALSE
.qualitative <- DirichletMultinomial:::.qualitative

pre_data_dmn <- L6_rela_fil_sAg_others[,unique(c(meta_fil_config$Previous_sra))]/1000
# pre_data0 <- L6_rela_fil_sAg_others[,unique(c(meta_fil_config[!meta_fil_config$Diease1 %in% c('CDI'), c('Previous_sra')]))]/100000
head(colSums(pre_data_dmn))
```

```
ERS1512354 ERS1512358 ERS1512361 ERS1512365 ERS1512369 ERS1512235
       100        100        100        100        100        100
```

Hide

```
pre_data_dmn_remove = noise.removal(pre_data_dmn, percent=0.01)
```

```
[1] 0.01
```

Hide

```
pre_data_dmn.dist=dist.JSD(pre_data_dmn_remove)

fit <- mclapply(1:7, dmn, count=t(pre_data_dmn_remove), verbose=TRUE)
```

```
dmn, k=1
```

```
  Soft kmeans
  Expectation Maximization setup
  Expectation Maximization
  Hessian
```

```
dmn, k=2
```

```
  Soft kmeans
    iteration 10 change 0.014519
    iteration 20 change 0.000243
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 0.073847
    iteration 20 change 0.000313
  Hessian
```

```
dmn, k=3
```

```
  Soft kmeans
    iteration 10 change 0.001769
    iteration 20 change 0.000027
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 11.248031
    iteration 20 change 0.304991
    iteration 30 change 0.002959
  Hessian
```

```
dmn, k=4
```

```
  Soft kmeans
    iteration 10 change 0.007000
    iteration 20 change 0.000435
    iteration 30 change 0.000016
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 9.836833
    iteration 20 change 0.256370
    iteration 30 change 0.000908
    iteration 40 change 0.000051
    iteration 50 change 0.000001
  Hessian
```

```
dmn, k=5
```

```
  Soft kmeans
    iteration 10 change 0.008515
    iteration 20 change 0.000070
    iteration 30 change 0.000001
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 3.843805
    iteration 20 change 0.150335
    iteration 30 change 0.461951
    iteration 40 change 0.302394
    iteration 50 change 0.013361
    iteration 60 change 0.000706
    iteration 70 change 0.000073
    iteration 80 change 0.000017
    iteration 90 change 0.000006
    iteration 100 change 0.000001
  Hessian
```

```
dmn, k=6
```

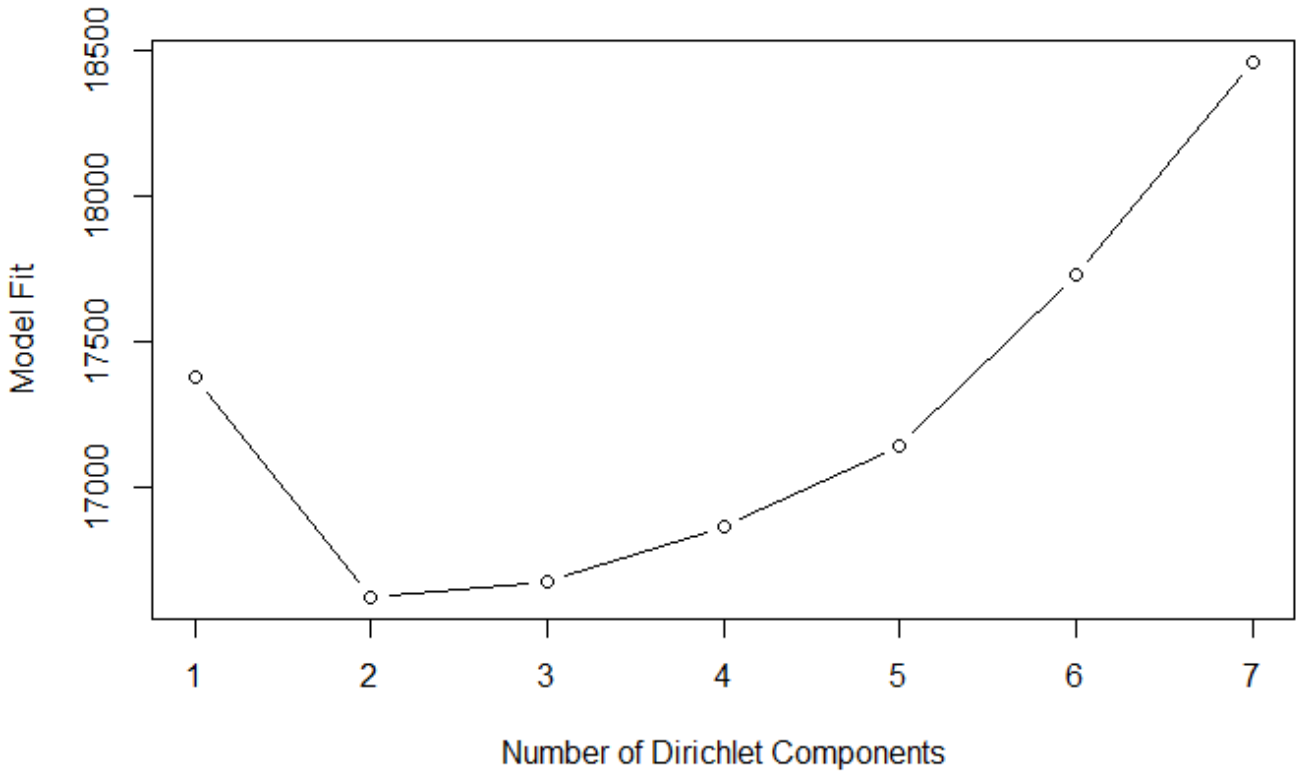
```
  Soft kmeans
    iteration 10 change 0.010590
    iteration 20 change 0.000763
    iteration 30 change 0.000361
    iteration 40 change 0.000221
    iteration 50 change 0.000158
    iteration 60 change 0.000126
    iteration 70 change 0.000111
    iteration 80 change 0.000105
    iteration 90 change 0.000108
    iteration 100 change 0.000119
    iteration 110 change 0.000145
    iteration 120 change 0.000195
    iteration 130 change 0.000306
    iteration 140 change 0.000604
    iteration 150 change 0.001887
    iteration 160 change 0.011249
    iteration 170 change 0.000805
    iteration 180 change 0.000023
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 9.717092
    iteration 20 change 0.846977
    iteration 30 change 0.411342
    iteration 40 change 0.075773
    iteration 50 change 0.000253
    iteration 60 change 0.000002
  Hessian
```

```
dmn, k=7
```

```
  Soft kmeans
    iteration 10 change 0.048999
    iteration 20 change 0.000293
    iteration 30 change 0.000002
  Expectation Maximization setup
  Expectation Maximization
    iteration 10 change 4.090679
    iteration 20 change 4.233082
    iteration 30 change 1.373933
    iteration 40 change 0.001562
    iteration 50 change 0.000005
  Hessian
```

Hide

```
lplc <- sapply(fit, laplace)
 # laplace goodness of fit
plot(lplc, type="b", xlab="Number of Dirichlet Components",
     ylab="Model Fit")
```



Hide

```
NA
NA
```

Hide

```
pdf(file='./figure1/lmain_dmn_class.pdf')
plot(lplc, type="b", xlab="Number of Dirichlet Components",
     ylab="Model Fit")
dev.off()
```

```
null device
          1
```

Hide

```
best <- fit[[which.min(unlist(lplc))]]
ass <- apply(mixture(best), 1, which.max)
```

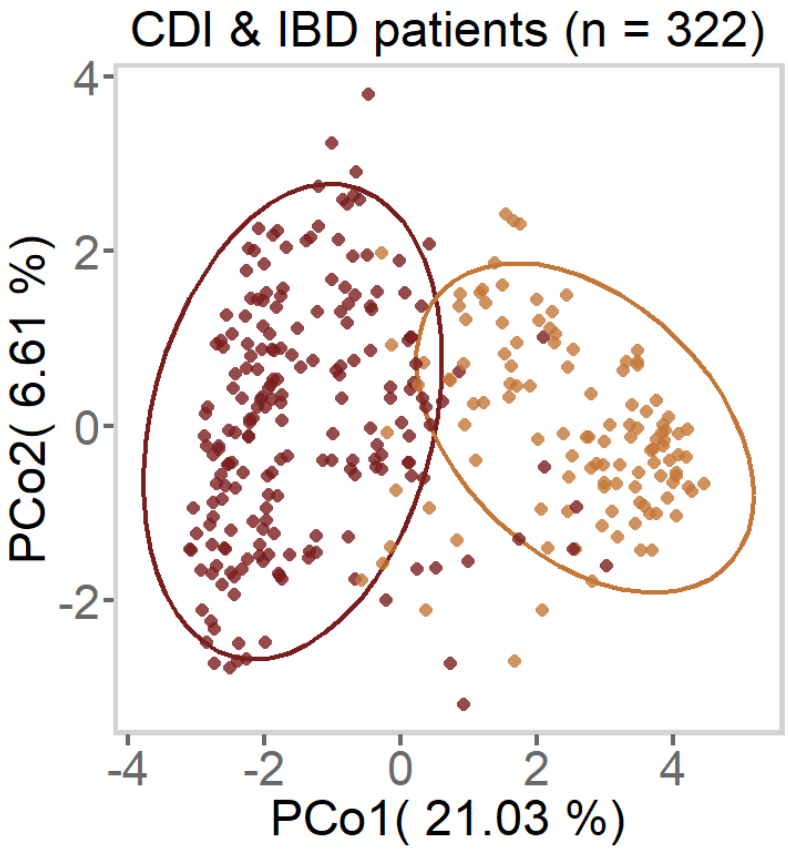
Hide

```
pre_obs.pcoa_dmn=dudi.pco(pre_data_dmn.dist, scannf=F, nf=2)
# s.class(pre_obs.pcoa$li, fac=as.factor(pre_data0.cluster), grid=F,sub="Principal coordiante a
nalysis")

##pcoa

PCo1 <- pre_obs.pcoa_dmn$li[ ,1]
PCo2 = pre_obs.pcoa_dmn$li[ ,2]

library(ggplot2)
# rownames(pre_obs.pcoa$li) == pre_sra_u[,"SRA"]

Groupn<-'postfmt'
# pre_sra_u[is.na(pre_sra_u[, Groupn]), Groupn]='NA'
# sample.groups = factor(pre_sra_u[,Groupn], levels=c('response', 'failure', 'NA'))
sample.groups <- 1
#Groupn<-'Diease'
#sample.groups = factor(pre_sra_u[,Groupn])

adonis(pre_data_dmn.dist ~ ass, permutations = 999)
```

```
Call:
adonis(formula = pre_data_dmn.dist ~ ass, permutations = 999)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

           Df SumsOfSqs MeanSqs F.Model     R2 Pr(>F)
ass         1    1188.7 1188.67  57.849 0.1531  0.001 ***
Residuals 320    6575.2   20.55         0.8469
Total     321    7763.9                 1.0000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
plotdata_dmn <- data.frame(rownames(pre_obs.pcoa_dmn$li),PCo1,PCo2, sample.groups, ass)
colnames(plotdata_dmn) <-c("sample","PCo1","PCo2","group", "data.cluster")
pc1 <-floor(pre_obs.pcoa_dmn$eig[1]*10000/sum(pre_obs.pcoa_dmn$eig))/100
pc2 <-floor(pre_obs.pcoa_dmn$eig[2]*10000/sum(pre_obs.pcoa_dmn$eig))/100
#sample.groups <- pre_sra_u[,Groupn]
#shape=factor(substr(pre_sra_u[,Groupn], 1, 1))

p<-ggplot(plotdata_dmn, alpha=I(0.8))+
  theme_classic()+
  stat_ellipse(aes(x=PCo1, y=PCo2, colour=as.factor(data.cluster),
                   group=as.factor(data.cluster)), level=0.9, size=1.5, show.legend = NA)+
  labs(title=paste("CDI & IBD patients (n = ", length(PCo1),")", sep=''), x=paste("PCo1(",pc1,
"%)"),y=paste("PCo2(",pc2,"%)") , colour="Cluster")+
  geom_point(aes(x=PCo1,y=PCo2, colour=as.factor(data.cluster), shape=factor(sample.groups),
                 alpha = factor(sample.groups)),size=4)+
  # theme(plot.title = element_text(size=21, hjust = 0.5),
  #     # title=element_text(family ="sans", size=21),
  #       text=element_text(family ="sans", size=18),aspect.ratio=0.95)+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust =
0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34
), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=1, legend.position = c(4, .65), legend.background=element_rect(fill = NA),
legend.text = element_text(size=18))+
  scale_colour_manual(values=c("#7A1D1E", "#C47737", "#E7A600"))+
  scale_alpha_manual(values = c(0.8))+
  theme(panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks = element_line(size=1.5, color
 ='dimgray'), axis.ticks.length = unit(7, "pt"));p
```



Hide

```
ggsave(paste("./figure1/1main_dmn", figli, ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
figli = figli + 1
```

Hide

```
library(coin)
# pre_obs.pcoa_dmn
pre_data_dmm.cluster <- paste('before', ass, sep = '')#beofre_entro[,2]
pre_data_dmm.prj <- beofre_entro[,3]

# options(stringsAsFactors = FALSE)
pre_data_dmm_remove <- pre_data_dmn_remove
pre_simp_name <- sapply(as.character(rownames(pre_data_dmm_remove)), simp_names)
pre_data_dmm_remove_dat <- data.frame(pre_data_dmm_remove, stringsAsFactors = F)
pre_data_dmm_remove_c <- as.data.frame(rbind(pre_data_dmm_remove_dat + 1e-5, pre_data_dmm.clust
er, pre_data_dmm.prj), stringsAsFactors =F)
rownames(pre_data_dmm_remove_c) <- c(pre_simp_name, 'pre_data_dmm.cluster', 'pre_data_dmm.prj')
t_pre_data_dmm_remove<-t(pre_data_dmm_remove_c)
t_pre_data_dmm_remove_c <- as.data.frame(t_pre_data_dmm_remove, stringsAsFactors = F)

t_pre_data_dmm_remove_c$pre_data_dmm.cluster <- as.factor(as.character(t_pre_data_dmm_remove_c
$pre_data_dmm.cluster))
t_pre_data_dmm_remove_c$pre_data_dmm.prj <- as.factor(t_pre_data_dmm_remove_c$pre_data_dmm.prj)

t_pre_data_dmm_remove_c$Enterobacteriaceae_ <- as.numeric(t_pre_data_dmm_remove_c$Enterobacteri
aceae_)
wilcox_test(Enterobacteriaceae_ ~ pre_data_dmm.cluster| pre_data_dmm.prj, t_pre_data_dmm_remove
_c)
```

```

    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Enterobacteriaceae_ by
     pre_data_dmm.cluster (before1, before2)
     stratified by pre_data_dmm.prj
Z = 8.8878, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

Hide

```
t_pre_data_dmm_remove_c$Bacteroides <- as.numeric(t_pre_data_dmm_remove_c$Bacteroides)
wilcox_test(Bacteroides ~ pre_data_dmm.cluster| pre_data_dmm.prj, t_pre_data_dmm_remove_c)
```

```
    Asymptotic Wilcoxon-Mann-Whitney Test

data:  Bacteroides by
     pre_data_dmm.cluster (before1, before2)
     stratified by pre_data_dmm.prj
Z = -9.3471, p-value < 2.2e-16
alternative hypothesis: true mu is not equal to 0
```

Hide

```
##sum genus to family for Enterobacteriaceae
keyword = 'f__Enterobacteriaceae'
genus_key <- sapply(as.character(grep(keyword, rownames(pre_data_dmm_remove), value = TRUE)), s
imp_names)
t_pre_data_dmm_remove_c_Entero <- colSums(apply(t_pre_data_dmm_remove_c[,genus_key], 1, as.nume
ric))
names(t_pre_data_dmm_remove_c_Entero) == rownames(t_pre_data_dmm_remove_c)
```

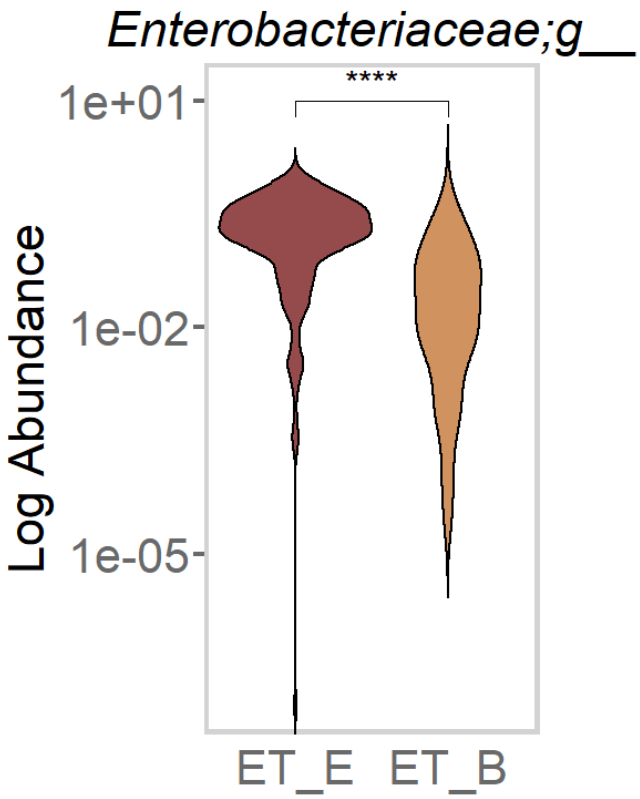
```
  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [26] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [51] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
 [76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[101] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[126] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[151] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[176] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[201] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[226] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[251] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[276] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[301] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE
```

Hide

```
plot_pre_data_dmm <- cbind((t_pre_data_dmm_remove_c_Entero)/100, t_pre_data_dmm_remove_c[,c('pr
e_data_dmm.cluster')], t_pre_data_dmm_remove_c[,c( 'Enterobacteriaceae_', 'Bacteroides')]/100)
colnames(plot_pre_data_dmm) <- c('Enterobacteriaceae', 'pre_data_dmm.cluster', 'Enterobacteriac
eae_', 'Bacteroides')
```

Hide

```
plot_pre_data_dmm1<- plot_pre_data_dmm
plot_pre_data_dmm1$pre_data_dmm.cluster <- ifelse(plot_pre_data_dmm1$pre_data_dmm.cluster == 'before1', 'ET_E', 'ET_B')
plot_pre_data_dmm1$pre_data_dmm.cluster <- factor(plot_pre_data_dmm1$pre_data_dmm.cluster , levels=c('ET_E', 'ET_B'))

p1<-ggviolin(plot_pre_data_dmm1, x="pre_data_dmm.cluster", y="Enterobacteriaceae_", fill = "pre_data_dmm.cluster",#fill = "", Enterobacteriaceae_
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab("Log Abundance")+labs(title='Enterobacteriaceae;g__')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme( aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p1
```



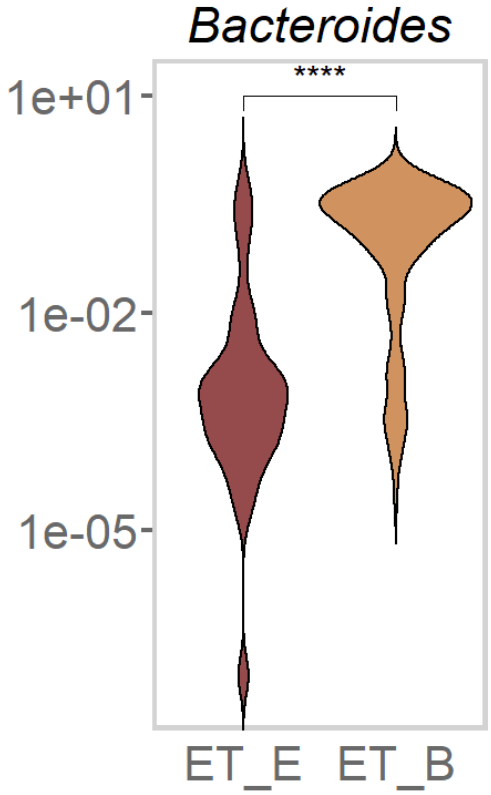
Hide

```
ggsave(paste("./figure1/fig_s1_dmm", 'figli1', ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
p2<-ggviolin(plot_pre_data_dmm1, x="pre_data_dmm.cluster", y="Bacteroides", fill = "pre_data_dmm.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Bacteroides')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5, face = 'italic'), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p2 #axis.title=element_text(size=21),
```



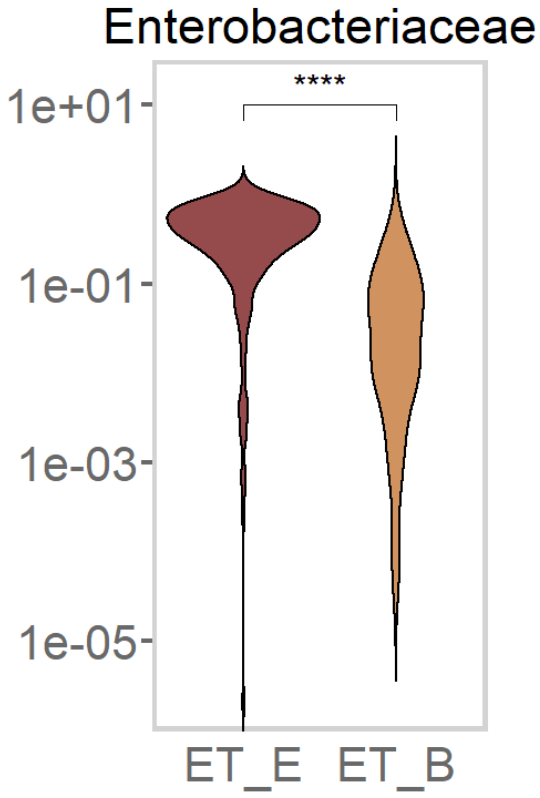
Hide

```
ggsave(paste("./figure1/fig_s1_dmm", 'figli2', ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
p3<-ggviolin(plot_pre_data_dmm1, x="pre_data_dmm.cluster", y="Enterobacteriaceae", fill = "pre_data_dmm.cluster",#fill = "",
             alpha = 0.8,  add.params = list(alpha=1), palette = c("#7A1D1E", "#C47737"), size = 0.8)+
  stat_compare_means(comparisons = list(c('ET_E', 'ET_B')), method = 'wilcox.test', label = "p.signif", label.x = 1.5, label.y = 1, size=8)+
  # yscale("log2", .format = FALSE)+
  theme_classic()+theme(legend.position = "right")+xlab(label = '')+ylab(" ")+labs(title='Enterobacteriaceae')+
  theme(text=element_text(family ="sans", size=32), plot.title = element_text(size=34, hjust = 0.5), axis.text = element_text(size=32, color ='dimgray'), axis.title.x = element_text(size=34), axis.title.y = element_text(size=34), axis.ticks = element_blank())+#, face = 'italic'
  theme(aspect.ratio=2,
        legend.direction = 'horizontal', legend.position = c(5, .15), legend.background=element_rect(fill = NA)
        ,panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
        ,axis.line=element_line(colour=NA, size = 0),axis.ticks.y = element_line(size=1.5, color ='dimgray'), axis.ticks.length = unit(7, "pt"))+
    scale_y_log10(expand = expansion(add=c(0, 0.5)));p3 #axis.title=element_text(size=21),
```



Hide

```
ggsave(paste("./figure1/fig_s1_dmm", 'figli3', ".pdf", sep = ''), device = "pdf")
```

```
Saving 12.9 x 8 in image
```

Hide

```
entero_dmn <- data.frame(sra = names(ass), dmm = ass)
entero_merged <- merge(merge(entero_dmn, beofre_entro, by.y='V1', by.x='sra'), beofre_entro_sep
rate, by.x='sra', by.y='names')
entero_merged$dmm <- ifelse(entero_merged$dmm %in% 1, 'before1', 'before2')

# lot_pre_data02_eb <- plot_pre_data02[order((3*as.numeric(plot_pre_data02$pre_data0.cluster) +
ifelse(plot_pre_data02$pre_data0.cluster == 'ET_E', 1-as.numeric(plot_pre_data02$Enterobacteria
ceae_), as.numeric(plot_pre_data02$Bacteroides)))),]
entero_merged <- entero_merged[order((3*as.numeric(entero_merged$V2) + ifelse(entero_merged$V2
 != entero_merged$sep & entero_merged$V2 != entero_merged$dmm, 2, ifelse(entero_merged$V2 != en
tero_merged$sep, 1, ifelse(entero_merged$V2 != entero_merged$dmm, 0.5,0))) )),]

entero_merged$sra <- factor(entero_merged$sra, levels = entero_merged$sra)
```

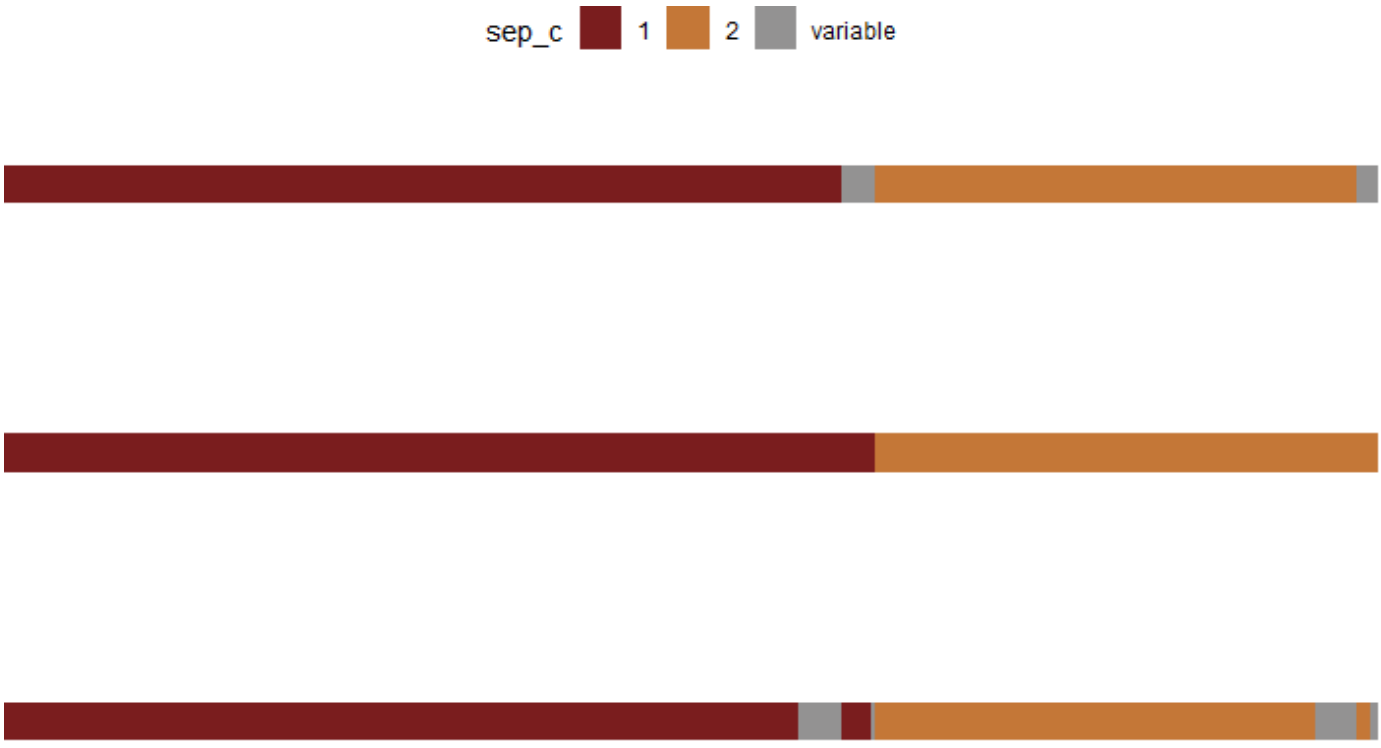
Hide

```
pp1<-ggplot(entero_merged)+
    geom_bar(mapping = aes(x = sra, y = 1, fill = sep_c),
       stat = "identity",
       width = 1)+
      theme_void()+
        scale_fill_manual(values=c("#7A1D1E", "#C47737", '#939292'))+
      theme(panel.spacing.x = unit(1, "mm"), aspect.ratio = 0.03)

pp2<-ggplot(entero_merged)+
    geom_bar(mapping = aes(x = sra, y = 1, fill = V2),
       stat = "identity",
       width = 1)+
      theme_void()+
    scale_fill_manual(values=c("#7A1D1E", "#C47737"))+
      theme(panel.spacing.x = unit(1, "mm"), aspect.ratio = 0.03)
pp3<-ggplot(entero_merged)+
    geom_bar(mapping = aes(x = sra, y = 1, fill = dmm_c),
       stat = "identity",
       width = 1)+
      theme_void()+
    scale_fill_manual(values=c("#7A1D1E", "#C47737", '#939292'))+
      theme(panel.spacing.x = unit(1, "mm"), aspect.ratio = 0.03)

ggpubr::ggarrange(pp1, pp2, pp3, ncol = 1, nrow = 3, common.legend = TRUE, heights = c(1, 1, 1
))
```

Hide

```
ggsave("./figure1/1main_unchanged.pdf", device = 'pdf')
```

```
Saving 7.29 x 4.5 in image
```



Hide

```
library(ggbeeswarm)
```

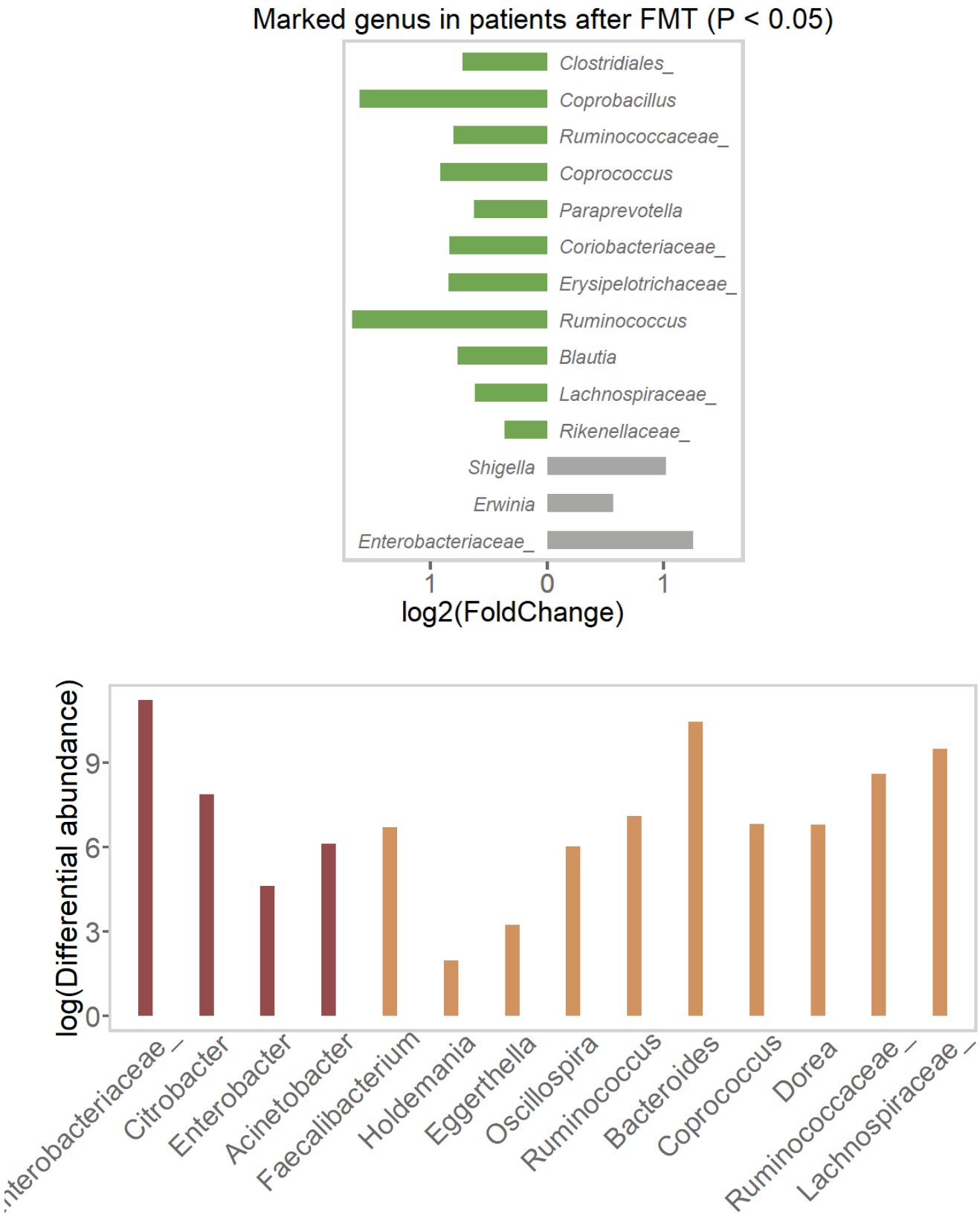
```
卯铃  愼狯卯牠昼狢愼钵愼狂ggbeeswarm愼钵愼貂挼狡挼犁撵铃挼铃R卯牠汾3.6.3 挼牠卯挝卯狯愼蚯撵挝玫狢卯牯挼挝
```

Hide

```
library(ggpubr)
library(coin)
library(reshape2)
###marked genus in response come from donor
#genus: L6_rela_fil_sAg_remove L6_rela_fil_sAg_remove_simp
#config: meta_fil_config1_na
#

# response_abundance <- function(pre_entro){
    ###marked genus in after response group
    # pre_entro<-'before2'
    tmp_after_response <- meta_fil_config1_na[meta_fil_config1_na$pre_entro %in% c(pre_entro),
 c('SRA_Sample', 'postfmt_symptoms', 'PRJ')]

    L6_rela_fil_sAg_remove_simp_after <- L6_rela_fil_sAg_remove_simp[,tmp_after_response$SRA_Sa
mple]

    # cols <- ncol(feature_abun_dat)
    group <- tmp_after_response$postfmt_symptoms
    prj <- tmp_after_response$PRJ
    seq(0.1, 0.9, 0.05) -> quan

    tmp_pval <- apply(L6_rela_fil_sAg_remove_simp_after, 1, function(x){
        pt <-  as.data.frame(cbind(x, group, prj))
        colnames(pt)<-c('nx', 'group', 'prj')
        # upt <- unique(pt)
        upt <- pt
        prj_list <- upt$prj
        list <- NULL
        for(i in unique(prj_list)){
            if (length(prj_list[prj_list %in% i]) > 2){
                list <- c(list, i)
            }
        }
        upt <- upt[upt$prj %in% list,]
        upt$nx <- as.numeric(as.character(upt$nx))
        tmp_test <- wilcox_test(nx ~ group | prj, upt)
        pval_a <- 1
        pval_a <- pvalue(tmp_test)
        if(is.na(pval_a)){pval_a<-1}
        a_nx <- upt[upt$group %in% c('response'), 'nx']
        p_nx <- upt[upt$group %in% c('failure'), 'nx']
        case <- quantile(log10(a_nx + 0.0001), quan)
        control <- quantile(log10(p_nx+ 0.0001), quan)
        gfc <- sum((case - control))/length(quan)
        return(c(pval_a, gfc))
    })

    ###select marked genus qvalue<0.05, and combine abundance
    tmp_pval_t <- data.frame(t(tmp_pval))
    colnames(tmp_pval_t) <- c('pval', 'gfc')
    # tmp_pval_t$id <- rownames(tmp_pval_t)
    # entero_diff_t$pval <- as.numeric(as.character(entero_diff_t$pval))
    tmp_qvalue <- p.adjust(tmp_pval_t$pval, method='fdr')
    tmp_pval_adjust <- cbind(tmp_pval_t, tmp_qvalue)
    tmp_pval_adjust05 <- tmp_pval_adjust[tmp_pval_adjust$tmp_qvalue < 0.05,]#tmp_qvalue < 0.0
5,]
```

```
    tmp_L6_05 <- L6_rela_fil_sAg_remove_simp[rownames(tmp_pval_adjust05),]
    tmp_L6_after_05 <- L6_rela_fil_sAg_remove_simp_after[rownames(tmp_pval_adjust05),]

    #tmp_L6_after_05  tmp_pval_adjust05
    #tmp_after_response
    mean_after_response <- apply(tmp_L6_after_05[,tmp_after_response$postfmt_symptoms %in% c('r
esponse')], 1, function(x){(mean(x))})#quantile((x + 0.0001), quan)
    mean_after_failure <- apply(tmp_L6_after_05[,tmp_after_response$postfmt_symptoms %in% c('fa
ilure')], 1, function(x){(mean(x))})
    plot_after_response <- as.data.frame(cbind(rownames(tmp_L6_after_05), as.numeric((mean_afte
r_response)), as.numeric((mean_after_failure))), stringsAsFactors = F)
    colnames(plot_after_response) <- c('genus', 'response', 'failure')
    plot_after_response$response <- (as.numeric(plot_after_response$response)+0)
    plot_after_response$failure <- (as.numeric(plot_after_response$failure)+0)
    plot_after_response$genus <- factor(plot_after_response$genus, levels = (rownames(tmp_pval_
adjust05)[order(sign(tmp_pval_adjust05$gfc)/tmp_pval_adjust05$tmp_qvalue, decreasing = F)]))
    plot_after_response$diff <- log2(plot_after_response$failure / plot_after_response$respons
e)#ifelse(plot_after_response$response > plot_after_response$failure, plot_after_response$respo
nse / plot_after_response$failure, plot_after_response$failure / plot_after_response$response)
    plot_after_response$qvalue <- tmp_pval_adjust05$tmp_qvalue
    min_y = 0
    mmin_y = -2

    pl<-ggplot(plot_after_response, aes(x = genus))+ #, aes(x = genus), color = sex
      geom_linerange(data = plot_after_response, aes(ymin = 0, ymax = diff, color=ifelse(respon
se > failure, "#4D9127", "#90908D")), size = 8, alpha = 0.8)+#ifelse(response > failure, mmin_
y, min_y)
      # geom_linerange(data = plot_after_response[plot_after_response$response < plot_after_res
ponse$failure,], aes(ymin = min_y, ymax = ), size = 6, alpha = 0.8, color=)+  -min_y+log10(qval
ue), min_y-log10(qvalue)
      geom_label(aes(x = genus, y = ifelse(response > failure, 0.1, -0.1), label = genus, famil
y = "sans",  hjust=ifelse((response > failure), 0, 1)),
             inherit.aes = F, fontface = "italic",
             size = 6, label.padding = unit(0.0, "lines"), label.size = 0,
             label.r = unit(0.0, "lines"), fill = "NA", alpha = 0.9, color = "dimgrey")+
      # scale_y_continuous(breaks = c(c(-2, -1, 0) - min_y, c(0, 1, 2, 3, 4, 5)+min_y),
                    # labels = c("2", "1", "0", "0", "1", "2", "3", "4", "5"))+
      # facet_wrap(~genus, ncol = 2)+
      scale_y_continuous(expand = expansion(mult=c(0.03, 0.15)), breaks = c(c(-2, -1, 0) - min_
y, c(0, 1, 2, 3, 4, 5)+min_y), labels = c("2", "1", "0", "0", "1", "2", "3", "4", "5"))+
      coord_flip()+
        labs(title="Marked genus in patients after FMT (P < 0.05)", x='', y="log2(FoldChange)",
colour="Cluster")+
      theme(text=element_text(family ="sans", size=24), plot.title = element_text(size=26, hjus
t = 0.5), axis.text = element_text(size=24, color ='dimgray'), axis.title.x = element_text(size
=26, hjust = 0.33), axis.title.y = element_text(size=26), axis.ticks = element_blank())+
      theme(aspect.ratio=1.3, legend.position = c(4, .65), legend.background=element_rect(fill
 = NA), legend.text = element_text(size=0))+
      scale_colour_manual(values=c("#4D9127", "#90908D", "#7A1D1E", "#C47737", "#E7A600"))+
      scale_alpha_manual(values = c(0.8))+
      theme(panel.background = element_rect(fill = NA, colour = "lightgrey", size = 3)
            ,axis.line=element_line(colour=NA, size = 0), axis.ticks.x = element_line(size=1.5,
color ='dimgray'), axis.ticks.length = unit(7, "pt")
            ,axis.text.y = element_text(size=0, angle = 0, color=NA)
            # ,axis.text.x = element_text(size=21, angle = 0)
            ,axis.ticks = element_blank())+theme(panel.grid.major = element_blank(), panel.gri
```

```
d.minor = element_blank())
    p1
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.